

Grup Seviyesinde Seyreklik için Normların Konveks Bileşimleri Üzerine On Convex Combinations of Norms For Group Sparsity

İlker Bayram

Elektronik ve Haberleşme Mühendisliği Bölümü,
İstanbul Teknik Üniversitesi
ilker.bayram@itu.edu.tr

ÖZETÇE

ℓ_1 normuna bağlı düzenlemeler, ‘seyrek’ işaretlerin onarımı, geri çatımı gibi problemlerde sıklıkla kullanılmaktadır. Fakat sadece seyrek olmayıp daha karmaşık yapılar barındıran işaretler söz konusu olduğunda ℓ_1 normu yeterli olamamaktadır. Bu gözlemden yola çıkarak, bu makalede ℓ_1 ve ℓ_2 normlarının konveks bileşimlerini inceliyoruz. Basit bir gürültü giderme düzenlemesinden de yararlanarak problemin çözümüne geometrik bir bakış getiriyoruz. Bu bakışın bahsi geçen normlara bağlı düzenlemelere ışık tutmasını umuyoruz.

ABSTRACT

Formulations based on the ℓ_1 norm are widely used for the restoration and/or reconstruction of ‘sparse’ signals. However, the ℓ_1 norm does not provide a rich enough model for signals that are not only sparse but also host more complex structures. Based on this observation, we investigate convex combinations of ℓ_1 and ℓ_2 norms. Employing a simple denoising formulation, we provide a geometric interpretation for the solution of the problem. We hope that this interpretation sheds light on formulations based on norms of the mentioned kind.

1. GİRİŞ

Gürültü giderme, işaret onarımı gibi problemlerde ℓ_1 normuna dayalı düzenlemeler oldukça sık kullanılmaktadır. Enerjinin sabit olduğu halde, ℓ_1 normunun küçük olması, işaretin ‘seyrekliği’ için uygun bir ölçü olmakla birlikte, karmaşık yapıdaki işaretler için yetersiz kalmaktadır. Bu makalede ℓ_1 ve ℓ_2 normlarının konveks bileşimlerinden oluşan norm ailesini inceleyeceğiz. Amacımız, bu norm ailesinin nasıl davrandığını basit bir düzenleme çerçevesinde daha iyi anlayabilmek.

Şöyle bir senaryo düşünelim : ‘ x ’ görüntülemek istediğimiz ‘seyrek’ nesne, ‘ y ’ de ‘ x ’e dair gürültülü gözlemlerimiz olsun. Bu durumda sıkça kullanılan bir düzenleme

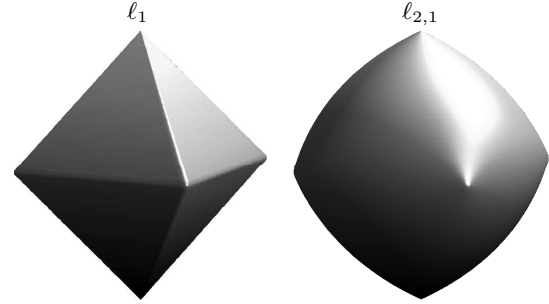
$$J(t) = \frac{1}{2} \|y - t\|_2^2 + \lambda \|t\|_1 \quad (1)$$

ceza fonksiyonunu kullanıp ‘ x ’,

$$\hat{x} = \underset{t}{\operatorname{argmin}} J(t) \quad (2)$$

şeklinde kestirmeyi önerir. Bu makalede $J(\cdot)$ fonksiyonunu

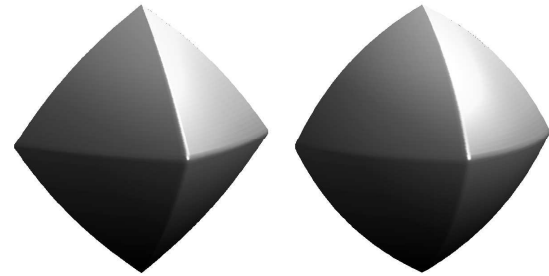
$$J_\alpha(t) = \frac{1}{2} \|y - t\|_2^2 + \lambda \left(\alpha \|t\|_1 + (1 - \alpha) \|t\|_2 \right) \quad (3)$$



ℓ_1 ve ℓ_2 Normlarının Konveks Kombinasyonları

$\alpha = 0.7$

$\alpha = 0.5$



Şekil. 1: Çeşitli normların birim topları. Üst panel : ℓ_1 normu ve (6)’da tanımlanan $\ell_{2,1}$ normu. Alt panel : $\alpha = 0.7$, ve $\alpha = 0.5$ için $\alpha \ell_1 + (1 - \alpha) \ell_2$ normları.

şeklinde değiştirip, kestirmeyi

$$\hat{x}_\alpha = \underset{t}{\operatorname{argmin}} J(t) \quad (4)$$

olarak yapmanın sonucu nasıl değiştireceğini inceleyeceğiz.

Bilindiği gibi ℓ_1 normuna dayalı (1) gibi ceza fonksiyonlarına bağlı kestirmeler seyrek olmaktadır¹. Bu, bazı problemler için arzulanan bir özellik olmasına rağmen, x_i ’lerin birbiriyle ilintili (correlated) olduğu durumlarda sonuçların gerçekten uzak olmasına da yol açabilmektedir. Bu gibi

¹Her ne kadar ‘seyrekliğin’ doğal tanımı ℓ_0 seyreklik sayısı (sparsity count) üzerinden yapılsa da bu fonksiyonel dışbükey olmadığı için, bu sayıdan türetilen düzenlemeler de dışbükey olmamaktadır. ℓ_1 normu, ℓ_0 sayısına bir anlamda ‘en yakın’ norm olmanın yanı sıra ℓ_1 normundan türetilen düzenlemelerin çözüm kümeleri, gerekli koşullar sağlandığında ℓ_0 sayısına dayanan düzenlemelerin çözüm kümelerine denktir.

durumlarda değişkenlerin bir çeşit komşuluk ilişkisine bağlı gruplar oluşturmalarına izin vermenin ve seyrekliği grup seviyesinde şart koşmanın daha uygun olduğu iddia edilmiştir [5]. Örneğin Kowalski ve Torrèsani, [5]'te ℓ_1 normu yerine karışık ('mixed') normlar kullanmayı önermişlerdir. $\ell_{2,1}$ karışık normu değişkenlerden $G_n = \{x_{n(1)}, x_{n(2)}, \dots, x_{n(N)}\}$ gibi 'K' tane grup oluşturulduğu halde

$$\|x\|_{2,1} = \sum_{n=1}^K \|G_n\|_2 \quad (5)$$

olarak tanımlanmıştır. Burada herbir grubun ℓ_2 normunun kullanılması, grup içerisinde seyreklik kısıtlamasını ortadan kaldırmaktadır. Buna basit bir örnek, üç değişkenli ($x = \{x_1, x_2, x_3\}$) bir durum için

$$\|x\|_{2,1} = \sqrt{x_1^2 + x_2^2} + \sqrt{x_1^2 + x_3^2} + \sqrt{x_2^2 + x_3^2} \quad (6)$$

şeklinde verilebilir. Bu normun (normalize edilmiş) birim topu Şekil 1'de gösterilmiştir. ℓ_1 normunun birim topuyla karşılaştırıldığında $\ell_{2,1}$ normunun korelasyonu düşük değişkenler üzerinde seyrekleştirici ('kutuplardaki köşelerin' sert olmasından dolayı), korelasyonu yüksek değişken gruplarında ise korelasyon yapısını muhafaza edici ('kutuplardan' uzakta ℓ_2 birim topuna benzerlik) özelliği bulunduğu söylenebilir. $\ell_{2,1}$ normunun bir dezavantajı seyreklik özelliğinin etkisini kontrol eden bir parametre bulunmamasıdır. Makalede inceleyeceğimiz ℓ_1 ve ℓ_2 normlarının konveks kombinasyonlarından oluşan norm ailesinin tanımındaki α parametresi tam da bu açıdan esnek bir yapı sağlamaktadır. Bu durum Şekil 1'de gösterilmeye çalışılmıştır. α parametresi sıfıra yaklaştıkça $\alpha \ell_1 + (1 - \alpha) \ell_2$ şeklinde tanımlanan normun birim topu ℓ_2 normunun birim topuna yakınsayacaktır.

Not 1 ℓ_1 ve ℓ_2 normlarının birim topları B_1 ve B_2 olarak ifade edildiği halde $\alpha \ell_1 + (1 - \alpha) \ell_2$ normunun birim topu $\alpha B_1 + (1 - \alpha) B_2$ değildir. Fakat bu tip bir ilişki normların 'dual' topları arasında bulunmaktadır. İzleyen bölümde bu ilişkiden faydalanarak problemin çözümünü bulmak için bir yöntem geliştireceğiz.

2. 'DUAL' PROBLEM

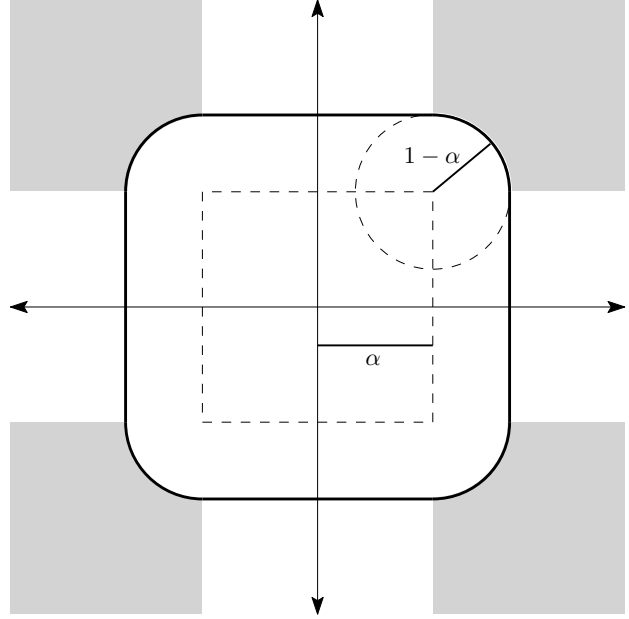
Tahmin edilebileceği gibi α parametresi sıfıra yaklaştıkça \hat{x}_α seyreklik özelliğini yitirecektir. Bunu daha net bir şekilde görebilmek için problemi biraz dönüştüreceğiz. Bunun için konveks analizden birkaç tanım ve sonuca ihtiyacımız olacak (konveks analiz hakkında kapsamlı bir kaynak için bkz. [4]).

Tanım 1 $C \subset \mathbb{R}^n$ dışbükey bir küme olsun. $x \in \mathbb{R}^n$ için $\sigma_C(x) = \sup_{z \in \mathbb{R}^n} \langle z, x \rangle$ şeklinde tanımlanan fonksiyona C kümesinin 'dayanak' (support) fonksiyonu denir.

\mathbb{R}^n üzerinde tanımlı herhangi bir normu uygun bir C kümesinin dayanak fonksiyonu olarak ifade edebiliriz. Örneğin, bu makalede ilgilendiğimiz ℓ_1 normu,

$$\|\cdot\|_1 = \sup_{|z_i| \leq 1} \langle z, \cdot \rangle = \sigma_{B_\infty}(\cdot) \quad (7)$$

şeklinde ifade edilebilir. Burada $B_\infty = \{z \in \mathbb{R}^n : |z_i| \leq 1\}$, ℓ_∞ normunun birim topunu temsil etmektedir. Benzer şekilde, B_2 , ℓ_2 normunun birim topu olduğu halde $\|\cdot\|_2 = \sigma_{B_2}(\cdot)$ denkliği de gösterilebilir.



Şekil 2: $\lambda = 1$ için ' $B^\alpha = \alpha B_\infty + (1 - \alpha) B_2$ ' kümesi. Küme, kenarları 2α uzunluğundaki bir kare üzerinde yarıçapı $1 - \alpha$ olan bir dairenin gezdirilmesiyle oluşturulabilir. B^α dışındaki beyaz alanlarda yer alan vektörlerin B^α 'ya izdüşümü seyrek olurken, gri alanlardaki vektörlerin izdüşümü seyrek değildir.

Normlara dayanak fonksiyonu olarak bakmak istememizin nedeni problemi basit bir şekilde çözebileceğimiz hale getirmemize yardımcı olmalarıdır. Bunun için iki sonuca ihtiyacımız olacak. Birincisi,

Önerme 1 C ve D , \mathbb{R}^n içinde dışbükey kümeler, λ da bir skaler olsun. Bu durumda aşağıdaki denklikler sağlanır.

- (a) $\lambda \sigma_C(\cdot) = \sigma_{\lambda C}(\cdot)$
- (b) $\sigma_C(\cdot) + \sigma_D(\cdot) = \sigma_{C+D}(\cdot)$

İkincisi ise,

Önerme 2 $C \subset \mathbb{R}^n$ kapalı dışbükey bir küme ve $P_C(\cdot)$, C kümesine izdüşüm operatörü olsun.

$$x = \underset{t}{\operatorname{argmin}} \frac{1}{2} \|y - t\|_2^2 + \sigma_C(t) \quad (8)$$

ise,

$$x = y - P_C(y) \quad (9)$$

eşitliği sağlanır.

Not 2 (a) Önerme 1, Tanım 1 kullanılarak gösterilebilir.

(b) Önerme 2 konveks analiz yöntemleriyle kısa bir kanıtla sahip olmakla birlikte (bkz. [2]), basit fakat biraz dolambaçlı bir şekilde de gösterilebilir. Yer darlığından dolayı bahsettiğimiz basit kanıtı veremiyoruz.

Önerme 1'i kullanarak (3)'teki $J_\alpha(\cdot)$ fonksiyonunu, K kümesi

$$K = \lambda(\alpha B_\infty + (1 - \alpha) B_2) \quad (10)$$

olarak tanımlandığı halde,

$$J_\alpha(t) = \frac{1}{2} \|y - t\|_2^2 + \sigma_K(t) \quad (11)$$

gibi yazabiliriz. Bu sayede, Önerme 2'den faydalanarak (4)'teki \hat{x}_α 'yı

$$\hat{x}_\alpha = y - P_K(y) \quad (12)$$

şeklinde niteleyebiliriz. Problemin çözümü için yapmamız gereken şey, $P_K(\cdot)$ operatörünü gerçekleştirmektir.

Öncelikle (10)'da tanımlanmış K kümesine daha yakından bakalım. Buradaki toplandan kasıt tam olarak,

$$K = \{z : \exists u/(\lambda\alpha) \in B_\infty, v/(\lambda(1-\alpha)) \in B_2, \text{ ki } z = u+v\}. \quad (13)$$

Bu küme $\lambda = 1$ olduğu halde \mathbb{R}^2 için Şekil 2'de gösterilmiştir. Bu kümeye B^α diyelim. Yine \mathbb{R}^2 üzerinden devam edersek, elimizde $y = \{y_1, y_2\}$ gibi bir vektör varsa yapmamız gereken y 'nin B^α 'ya izdüşümünü bulmaktır. P_y olarak ifade edeceğimiz bu izdüşümü kolayca elde edebiliriz :

- (1) y noktasının αB_∞ üzerine izdüşümünü bulup buna p_∞ diyelim.
- (2) $(y - p_\infty)$ noktasının $(1 - \alpha)B_2$ üzerine izdüşümünü bulup buna p_2 diyelim.
- (3) $P_y = p_1 + p_2$.

Bu yöntem Şekil 3'te gösterilmiştir.

Not 3 (a) Yöntemin gerçekten de izdüşüm noktasını bulduğu $y - P_y$ vektörünün B^α kümesinin P_y noktasındaki tanjantına (ki bu durumda tek bir tanjant vardır) dik olduğu gösterilerek yapılabilir.

(b) Genelde iki küme toplamına izdüşümü hesaplamak için, tartıştığımız yöntemin adımlarını tekrarlamak gerekmektedir (bkz. [1]). Fakat buradaki özel durumda yinelemeye ihtiyaç kalmamıştır.

Bu yöntemden faydalanarak, (4)'teki problemin çözümünü (ki $(y - P_y)$ olduğunu biliyoruz) şu şekilde çıkarabiliriz.

Önerme 3 $J_\alpha(\cdot)$ ve \hat{x}_α , (3) ve (4)'teki gibi tanımlanmış olsun.

$$z = \text{soft}(y, \lambda\alpha) \quad (14)$$

$$= \text{sign}(y) \cdot \max\{|y| - \lambda\alpha, 0\} \quad (15)$$

olduğu halde

$$\hat{x}_\alpha = \frac{z}{\|z\|_2} \cdot \max\{0, (\|z\|_2 - \lambda(1-\alpha))\} \quad (16)$$

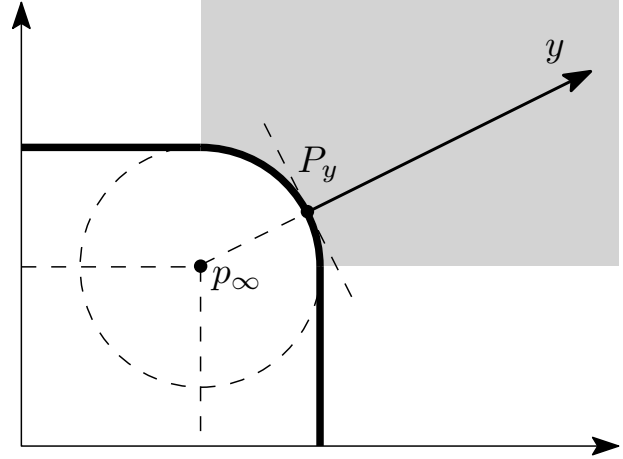
denkliği sağlanır.

' α ' Parametresinin Seyreklik Üzerine Etkisi

Önerme 3'e göre \hat{x}_α 'yı hesaplarken ilk adımda uygulanan 'soft thresholding' işleminin seyrekleştirici bir etkisi vardır. \mathbb{R}^2 'yi düşünecek olursak, y_1 ve y_2 gözlemlerinin birbiriyle ilintisinin 'yüksek olmadığı' durumlarda (Şekil 2'de B^α 'nın dışında kalan beyaz alan) çözüm vektörünün sadece bir tane sıfırdan-farklı bileşeni olacaktır. y_1 ve y_2 'nin birbiriyle ilintisinin 'yüksek olduğu' durumlarda ise (Şekil 2'de B^α 'nın dışında kalan gri alan) çözüm vektörünün her iki bileşeni de sıfırdan farklı olacaktır. α parametresi, ilintinin tam da ne zaman yüksek sayılacağını belirler (bkz. Şekil 2).

Problemin 'Elastik Ağ' ile İlişkisi

Makalede incelediğimiz (3)'teki ceza fonksiyonu, Zou ve Hastie'nin önerdiği 'Elastic Net' ceza fonksiyonuyla yakından



Şekil 3: y noktasının Şekil 2'deki kümeye izdüşümünün bulunması. $y - P_y$ vektörü, kümenin P_y noktasından geçen tanjantına diktir.

ilişkilidir (bkz. [7]). Elastic Net ceza fonksiyonu

$$L(\lambda_1, \lambda_2, t) = \|y - At\|^2 + \lambda_1 \|t\|_1 + \lambda_2 \|t\|_2^2 \quad (17)$$

şeklinindedir. $L(\lambda_1, \lambda_2, t)$ ceza fonksiyonunu minimize eden \hat{t} değeri ise $A = I$ için (ki yazarlar tarafından 'naive elastic net estimate' olarak adlandırılmıştır)

$$\hat{t} = \frac{\text{soft}(y, \lambda_1/2)}{1 + \lambda_2} \quad (18)$$

olarak bulunmuştur.

Elastic Net problemi uygun λ_1 ve λ_2 değerleri için $A = I$ olduğu durumda (3)'teki ceza fonksiyonunu minimize etmeye denktir. Genelde, bahsettiğimiz uygun λ_1, λ_2 değerlerini bulmak kolay olmasa da bu durumda (18) ve Önerme 3'ü karşılaştırarak

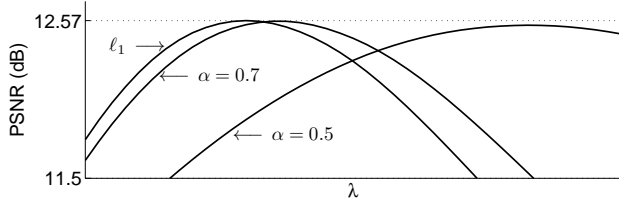
$$\lambda_1 = 2\lambda\alpha, \quad \lambda_2 = \frac{\lambda(1-\alpha)}{\|\text{soft}(y, \lambda\alpha)\| - \lambda(1-\alpha)} \quad (19)$$

olduğu görülebilir. Bu makaledeki incelememizin bu denklitten yararlanarak 'Elastic Net' temelli yöntemlere de farklı bir bakış açısı getireceğini umuyoruz.

3. DENEYLER VE TARTIŞMA

Şekil 5(a)'da tipik bir kısmı gösterilen bir işaret düşünelim². Bu işaretin az sayıdaki sıfırdan farklı örneğinin kümeler halinde öbeklendiğini, dolayısıyla 'Giriş'te tartıştığımız modele uyduğunu söyleyebiliriz. Bu işarete bir miktar gürültü ekleyerek elde edilen, Şekil 5(b)'de gösterilen işaret de 'Gözlem' işaretimiz olsun (yani (3)'teki y işareti - SNR = 10dB). Bu gözlem işaretini kullanarak çeşitli α değerleri için (4)'teki \hat{x}_α 'yı hesapladık. Şekil 4'te değişen SNR'ın farklı α değerleri için λ 'ya bağlı değişimi gösterilmektedir. Genel olarak, α değeri azaldıkça ($\alpha = 1$, ℓ_1 normunu vermektedir) SNR eğrisinin yumuşayarak ötelendiğini ve en yüksek SNR değerinin fazla olmasa da bir miktar düştüğünü gözlemledik. En yüksek SNR değerinin α arttıkça düşmesi her ne kadar beklemediğimiz

²İşaret, bir konuşma işaretinin rasyonel genleşmeli dalgacık altbantlarından birisidir.



Şekil 4: Farklı α değerleri için SNR'ın λ 'ya bağlı değişimi. Soldan sağa, sırayla $\alpha = 1$ (yani ℓ_1), $\alpha = 0.7$ ve $\alpha = 0.5$ seçimlerinin verdiği eğriler gösterilmiştir.

(ve arzu etmediğimiz) bir sonuç olsa da, eğrinin yumuşayarak λ 'ya daha az bağımlı hale gelmesinin de ilginç olduğunu düşünüyoruz. Bu tip problemlerde en iyi λ 'nın seçilmesinin her zaman mümkün olmadığı göz önünde bulundurulursa bu özelliğin önem kazanabileceğini söyleyebiliriz.

Değişen α değerleri için \hat{x}_α 'nın davranışı ise beklediğimiz yönde gerçekleşmiştir. $\alpha = 1$ (ℓ_1 normu) ve $\alpha = 0.5$ için bulunan \hat{x}_α 'lar Şekil 5(c) ve (d)'de görülebilir. Bu iki işaret orijinal işareti kestirmede farklı açılardan başarılı olmuştur. ℓ_1 normuna bağlı işaret, orijinal işaretin sıfıra yakın kesimlerini daha iyi belirlerken $0.5\ell_1 + 0.5\ell_2$ normuna bağlı işaret orijinal işaretin sıfırdan farklı olduğu bölgelerde daha iyi bir kestirim yapmıştır. Bu fark Şekil 5(e)'de daha açık bir şekilde görülebilir.

4. SONUÇ

Bu çalışmada işaret işleme seyreklik düzenlemesinin başarımını iyileştirmek için ℓ_1 ve ℓ_2 normlarının konveks bileşimlerini inceledik. Basit bir 'gürültü giderme' formülasyonunun 'dual' probleminin basit bir geometrik yorum ve gerçekleşmesi kolay, tek adımlık bir çözüme sahip olduğunu gösterdik. Bu çözümün problemin yakını ilişkisi içinde bulunduğu 'Elastic Net' problemini de anlamakta fayda sağlayacağını düşünüyoruz. İncelediğimiz problem her ne kadar 'gürültü giderme' düzenlemesine sahip olsa da,

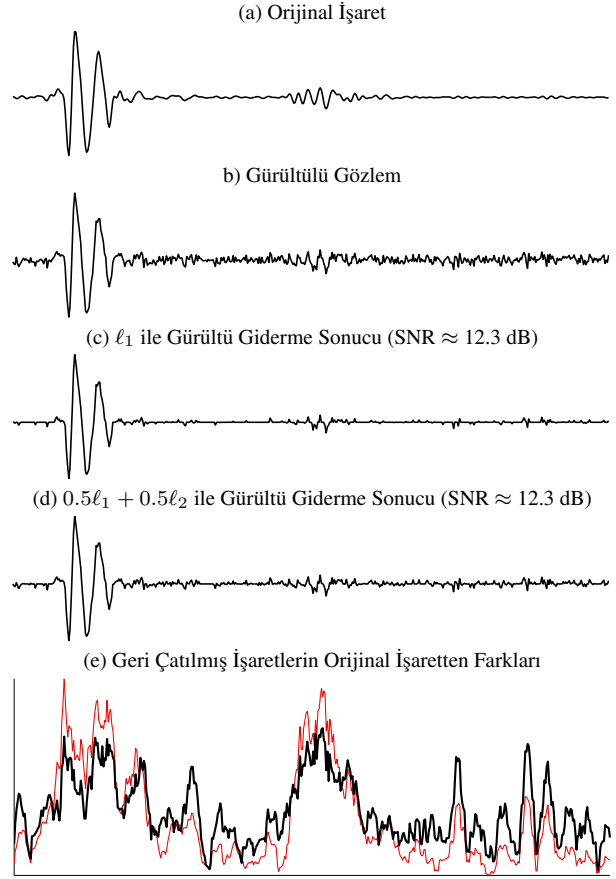
$$\tilde{J}_\alpha(t) = \frac{1}{2} \|y - A, t\|_2^2 + \lambda \left(\alpha \|t\|_1 + (1 - \alpha) \|t\|_2 \right) \quad (20)$$

tipi daha genel düzenlemelerde de 'Majorization Minimization' gibi teknikler sayesinde (bkz. [3]) iteratif algoritmaların içerisine de yerleştirilebilir.

Makalede ele almadığımız bir konu ise α ve λ parametrelerinin seçimi ile ilgiliydi. Elimizde bu parametrelere bağlı tek adımlık bir fonksiyon olduğunu da göz önünde bulundurduğumuzda 'Stein's Unbiased Risk Estimator' (SURE) temelli (bkz. [6]) yöntemlerin kullanılabilirliğini söyleyebiliriz. Bu problemi yakın gelecekte ele almayı planlıyoruz.

5. KAYNAKÇA

- [1] İ. Bayram. Mixed norms with overlapping groups as signal priors, ICASSP 2011.
- [2] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, Ocak-Mart 2004.
- [3] M. A. T. Figueiredo, J. M. Bioucas-Dias, ve R. D. Nowak. Majorization-minimization algorithms for wavelet-based



Şekil 5: (a) Orijinal işaret bir konuşma işaretinin rasyonel genleşmeli dalgacık dönüşümünün altbantlarından birisidir. (b) Gözlem işaretini (a)'daki işarete beyaz bağımsız Gaus gürültüsü ekleyerek elde ettik (SNR=10dB). (c) $\alpha = 1$ (ℓ_1 normu) için \hat{x}_α . (d) $\alpha = 0.5$ için \hat{x}_α . (e) Gösterilen \hat{x}_α 'ların orijinal işareten farklarının mutlak değeri. Burada ince kırmızı $\alpha = 1$, kalın siyah $\alpha = 0.5$. Bu şekil yapılan kestirimin α parametresine bağlı olarak nasıl davrandığını göstermektedir – α azaldıkça kestirimin başarımı, seyrekliğin yüksek olduğu yerlerde düşmekte, sıfırdan farklı örnek öbeklerinde yükselmektedir.

image restoration. *IEEE Trans. Image Processing*, 16(12):2980–2991, Aralık 2007.

- [4] J.-B. Hiriart-Urruty ve C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2001.
- [5] M. Kowalski ve B. Torrèani. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. *Signal, Image and Video Processing*, 3(3):251–264, Eylül 2009.
- [6] F. Luisier, T. Blu, ve M. Unser. A new SURE approach to image denoising. *IEEE Trans. Image Processing*, 16(3):593–606, Mart 2007.
- [7] H. Zou ve T. Hastie. Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*, 67(2):301–320, Nisan 2005.